MATH 6380Q Project 1 Peer Review

03. LUI, Go Nam. Principle Component Analysis on Finance Data

[ Reviewer 1 ]

1. Principle Component Analysis on Finance Data

Summary:

To analyze the time series data of the finance data from SNP500 dataset, the author first divided the stocks into 10 different groups based on industry. For each group, PCA followed by parallel analysis were performed to find the principle components which explained most of the variations in the time series data. Lastly, the data projected was by the first principle component in each group and compared amongst each other.

Strength of the project:

The report is clean and concise, the method the author used was simple yet reasonable to identify the trend of the industries.

Weakness of the project:

In the figures, only the magnitude of the eigenvalues of the decomposition were displayed. However, the percentage variance explained by each principle components is not shown. As a result, the decision of picking only the first PC being representable enough, is questionable.

|  |  |
| --- | --- |
| Evaluation on Clarity and quality of writing (1-5): | 4 |
| Evaluation on Technical Quality (1-5):  | 4 |
| Overall rating:  | 4 |
| Confidence on your assessment: | 1 |

[ Reviewer 2 ]

In poster three, “Principle Component Analysis on Finance Data”, author used PCA method to discover the development trend for different industry based on the SNP500 data. For each industry group, author get their first principle component and use it for the inter-industries comparison. Result is financial industry has the rapidest increasing speed. Other increasing industries includes industrials, health care and materials etc. The conclusion is using principle component analysis is acceptable for the time series stock price and by applying this technology, use one principle component to represent the whole industry group, to check the trend and investigate the relationship of different industry.

The strength of this poster is the author's ideas are clear, and the background, solutions, results and analysis of the questions are raised. The author successfully uses PCA to analyze the stock. The results are technically sound.

The weakness of this poster is lack of mathematical language and content can be further enriched.

Evaluation on Clarity and quality of writing: 3. The report is clearly written and there is a good use of examples and figures. Suggestion is to further enrich the content and add mathematical language.

Evaluation on Technical Quality: 4.

Overall rating: 3.

Confidence on my assessment: 3.

[ Reviewer 3 ]

03. LUI, Go Nam. Principle Component Analysis on Finance Data

This poster is mainly intended to discuss how many principle components could be used to described the evolving trend of the closed price. It was revealed that the first principle component is enough to do so.

One good point is that the author divided the whole SNP500 dataset into 10 categories which is very important in the further analysis process.

However, labelling the PCA transformed data as Hedonic closed price is not so proper in the figure of Section 4 in that closed price cannot be negative.

The poster has a simple organization. In the detailed analysis, the author did not refer to any example which can be important in the evolving process in the stock market. There are some grammar error in Section 5, Line 8: ‘includes’ should be replaced by ‘include’; Line 10: ‘have’ should be changed to ‘having’.

This project does not cover too many technical issues. Therefore, the result is very easy to follow. The result is very obvious to observe and the reasoning is solid. The author did not touch on strength/weakness evaluation of the method adopted. For reference, papers on PCA method should be cited.

It can be rated as 3 and my confidence is 2.

[ Reviewer 4 ]

This report is well organized and easy to understand. The figures are good and lots of explanation are given.

Evaluation on Clarity and quality of writing: 5

Evaluation on Technical Quality:3

Overall rating:3

Confidence on your assessment:2

[ Reviewer 5 ]

Summary: PCA is used to discover the development trend for different industry based on the SNP500 data.

Strengths: Different sectors are considered.

Weakness: Different sectors are only compared using time series without PCA analysis.

Evaluation on Clarity and quality of writing: 4.

Evaluation on Technical Quality: 3.

Overall rating: 3

Confidence on your assessment: 2

[ Reviewer 7 ]

**3. Principle Component Analysis on Finance Data**

* **Summary of this report:** The PCA is used to analyze the financial data. The principal components are shown in this report. The main conclusion in this report is that the first principle component can largely explain the price trend in the data.
* **Describe the strengths of the report:** The PCA is correctly used and the top 12 principle components are shown.
* **Describe the weaknesses of the report:** The discussions in this poster is a little few. The main conclusion is that the first principle component can represent the whole industry group. However, further investigations on the principle components should be conducted.
* **Evaluation on Clarity and quality of writing (1-5): 3**

Details of typos: 1)  Other increasing industries includes (include) industrials, health care and materials etc. 2) This dataset includes
the closed price of work days in four year (years) for 452 different company (companies). 3) Also, more variables for the companies should be considered, such as the location of the company, to construct ~~a~~ more thorough research.

* **Evaluation on Technical Quality (1-5): 3**

The discussions and the data analysis are insufficient.

* **Overal ratings: 3**
* **Confidence on your assessment: 3**

 [ Reviewer 8 ]

• Summary of the report.

This report applies PCA on financial data.

• Describe the strengths of the report.

Clearly written.

• Describe the weaknesses of the report.

Lack of deeper analysis. The methods (just PCA) might be a little bit too simple.

• Evaluation on Clarity and quality of writing (1-5): Is the report clearly written? Is there a good use of examples and ﬁgures? Is it well organized? Are there problems with style and grammar? Are there issues with typos, formatting, references, etc.? Please make suggestions to improve the clarity of the paper, and provide details of typos.

4

• Evaluation on Technical Quality (1-5): Are the results technically sound? Are there obvious ﬂaws in the reasoning? Are claims well-supported by theoretical analysis or experimental results? Are the experiments well thought out and convincing? Will it be possible for other researchers to replicate these results? Is the evaluation appropriate? Did the authors clearly assess both the strengths and weaknesses of their approach? Are relevant papers cited, discussed, and compared to the presented work?

2

• Overall rating: (5- My vote as the best-report. 4- A good report. 3- An average one. 2below average. 1- a poorly written one).

3-

• Conﬁdence on your assessment (1-3) (3- I have carefully read the paper and checked the results, 2- I just browse the paper without checking the details, 1- My assessment can be wrong)

2+

[ Reviewer 9 ]

1. **Summary**

The author uses PCA and parallel analysis to find the development trend for the different industries based on the SNP500 data. The data is divided into 10 groups and the inter-industries comparison is done based on the first principal component.

1. **Strength of the report**

The data trend is well described using the PCA analysis. The conclusions such as “most of the industries are increasingly developing in this 4 years, and financial industry has the rapidest increasing speed. Other increasing industries includes industrials, health care and materials etc.” is of value to this data analysis.

1. **Weakness of the report**

The details of the method PCA is not provided and there lack of a specific number in the analysis.

1. **Evaluation of clarity and quality of writing**

The report is clearly written and well organized. The figures are clearly explained. Small typos such as in Introduction part: “different industry” should be “different industries”.

1. **Evaluation on technical quality**

The results are technically sound and no obvious flaws in the reasoning are found. The report provides the methods for analyzing similar dataset. However, the author gives the strength of using PCA to do the analysis, and also provide possibilities to use other method to do the data analysis.

1. **Overall rating**

An overall rating of 4 is given.

1. **Confidence on your assessment**

2- I just browse the paper without checking the details

[ Reviewer 10 ]

*Summary:*

Conducted PCA on the price of SNP500.

*Strength:*

Did parallel analysis to ensure the result from PCA.

*Weakness:*

Operate PCA directly on the original price data.

*Evaluation on Clarity and quality of writing (1-5):2*

The section ‘data rearrangement’ sounds like some preprocess or transformation of data for further analysis, but the writer present a histogram of company numbers by different industry sectors, which is more narrow than ‘rearrangement’. Maybe he/she can use more accurate description for this section.

The section 4 of Comparison by PCA has only one figure, it’s difficult for reader to guess if it’s the figure discussed at section 5.

*Evaluation on Technical Quality (1-5):2*

The distribution of PCA input is important. If the stock price is preprocessing into percentage changes in stock price, the parameter is nearly normal distributed. The original stock price data is more random and may have different scale when the company is different.

*Overall rating (1-5): 2*

*Conﬁdence on your assessment (1-3): 3*

[ Reviewer 11 ]

Summary: The author does PCA on SNP500 dataset and use parallel analysis to determine the components.

Strengths: Clearly states the method and results.

Weaknesses: The report just do PCA on the dataset and do not have any deeper analysis.

Writing: (5) The report is well-written and well-organized.

Technical Quality: (3) There is an obvious flow in the reasoning. But it seems that the author just plots the stock closed prices using first principle component. I think the author can do more deeper analysis.

Overall rating: (3)

Confidence on my assessment: (2)

[ Reviewer 12 ]

**Summary:** This project uses PCA to analyze the SNP 500 dataset and especially the finance data as a subset of the whole dataset. By dividing all the stocks into 10 clusters, this project implements PCA on each cluster and uses the first principal component score to show the trend of each cluster’s stock price over time.

**Strengths:** This project clearly divides all the stocks into 10 groups and uses PCA on each cluster. The results are clear to get.

**Weaknesses:** The figures are not very clear. Explained variance ratio is a better choice to plot than the true eigenvalues. We cannot get a direct conclusion that the projection values on the first principal component direction reflect the overall trend of the price of a specific cluster. It’s better to compare this with the raw data to see the meaning of the projection values. Besides, there are some grammatical mistakes and the sentences are too informal.

**Evaluation on Clarity and quality of writing:** 3

**Evaluation on Technical Quality:** 3

**Overall rating:** 3

**Confidence on your assessment:** 2